Load the libraries

```
library(caret)
library(pander)
library(dplyr)
```

```
data("GermanCredit")
```

• Build a regression model to predict variable "Amount" as a function of other variables used following methodology:

1. Split sample randomly into training-test using a 632:368 ratio.
2. Build the model using the 63.2% training data and compute R-square in holdout data. (function lm() gives R-squares)
3. Save the coefficients, R-square in training and holdout samples. (To compute R-square in Holdout, take the square of correlation between actual and predicted values)

We'll first select best predictor variables based on step-wise regressions

```
# fit a model
fit <- lm(Amount~., data = GermanCredit)
# fit stepwise model
step_fit <- step(fit, trace = 0)
summary(step_fit)
```

```
##
## Call:
## lm(formula = Amount ~ Duration + InstallmentRatePercentage +
##     Telephone + Class + CheckingAccountStatus.lt.0 + CheckingAccountStatus.gt.200 +
##     CreditHistory.NoCredit.AllPaid + Purpose.NewCar + Purpose.UsedCar +
##     Purpose.Furniture.Equipment + Purpose.Radio.Television +
##     Purpose.DomesticAppliance + Purpose.Repairs + Purpose.Education +
##     Purpose.Retraining + Purpose.Business + SavingsAccountBonds.lt.100 +
##     SavingsAccountBonds.100.to.500 + SavingsAccountBonds.500.to.1000 +
##     EmploymentDuration.gt.7 + Personal.Male.Single + OtherDebtorsGuarantors.CoApplicant +
##     Property.RealEstate + Property.Insurance + Property.CarOther +
##     Job.UnemployedUnskilled + Job.UnskilledResident + Job.SkilledEmployee,
##     data = GermanCredit)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5254.8 -1048.2  -125.2   672.9 10961.9
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      7026.770    623.306  11.273  < 2e-16
## Duration                          125.208      5.336  23.464  < 2e-16
## InstallmentRatePercentage        -802.738     53.473 -15.012  < 2e-16
## Telephone                        -499.824    130.916  -3.818 0.000143
## ClassGood                        -353.331    138.654  -2.548 0.010978
## CheckingAccountStatus.lt.0       -285.175    140.483  -2.030 0.042633
## CheckingAccountStatus.gt.200     -672.257    240.038  -2.801 0.005202
## CreditHistory.NoCredit.AllPaid    829.486    302.837   2.739 0.006275
## Purpose.NewCar                  -1741.606    553.753  -3.145 0.001711
## Purpose.UsedCar                 -1084.157    564.176  -1.922 0.054940
```

```
## Purpose.Furniture.Equipment        -1783.742    559.441  -3.188 0.001476
## Purpose.Radio.Television           -2032.631    555.494  -3.659 0.000267
## Purpose.DomesticAppliance          -2388.669    759.651  -3.144 0.001715
## Purpose.Repairs                    -1663.769    667.728  -2.492 0.012880
## Purpose.Education                  -1886.952    601.819  -3.135 0.001768
## Purpose.Retraining                 -2205.528    818.799  -2.694 0.007190
## Purpose.Business                   -1926.931    569.487  -3.384 0.000744
## SavingsAccountBonds.lt.100          -287.400    147.047  -1.954 0.050931
## SavingsAccountBonds.100.to.500      -517.519    218.107  -2.373 0.017849
## SavingsAccountBonds.500.to.1000     -623.151    258.442  -2.411 0.016086
## EmploymentDuration.gt.7             -218.619    137.967  -1.585 0.113390
## Personal.Male.Single                 487.963    120.636   4.045 5.65e-05
## OtherDebtorsGuarantors.CoApplicant   618.368    294.898   2.097 0.036262
## Property.RealEstate                 -700.996    205.697  -3.408 0.000682
## Property.Insurance                  -444.779    200.521  -2.218 0.026778
## Property.CarOther                   -444.190    185.526  -2.394 0.016845
## Job.UnemployedUnskilled            -1683.717    430.727  -3.909 9.91e-05
## Job.UnskilledResident              -1245.189    227.671  -5.469 5.75e-08
## Job.SkilledEmployee                -1290.869    182.932  -7.057 3.24e-12
##
## (Intercept)                        ***
## Duration                           ***
## InstallmentRatePercentage          ***
## Telephone                          ***
## ClassGood                          *
## CheckingAccountStatus.lt.0         *
## CheckingAccountStatus.gt.200       **
## CreditHistory.NoCredit.AllPaid     **
## Purpose.NewCar                     **
## Purpose.UsedCar                    .
## Purpose.Furniture.Equipment        **
## Purpose.Radio.Television           ***
## Purpose.DomesticAppliance          **
## Purpose.Repairs                    *
## Purpose.Education                  **
## Purpose.Retraining                 **
## Purpose.Business                   ***
## SavingsAccountBonds.lt.100         .
## SavingsAccountBonds.100.to.500     *
## SavingsAccountBonds.500.to.1000    *
## EmploymentDuration.gt.7
## Personal.Male.Single               ***
## OtherDebtorsGuarantors.CoApplicant *
## Property.RealEstate                ***
## Property.Insurance                 *
## Property.CarOther                  *
## Job.UnemployedUnskilled            ***
## Job.UnskilledResident              ***
## Job.SkilledEmployee                ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1798 on 971 degrees of freedom
## Multiple R-squared:  0.6058, Adjusted R-squared:  0.5944
```

```
## F-statistic:  53.3 on 28 and 971 DF,  p-value: < 2.2e-16
```

Based on step-wise regression we've selected a final model with 60.58% accuracy with the following variables as predictor variable.

Duration,InstallmentRatePercentage, Telephone,Class,CheckingAccountStatus.lt.0,CheckingAccountStatus.gt.200, CreditHistory.NoCredit.AllPaid,Purpose.NewCar,Purpose.UsedCar, Purpose.Furniture.Equipment,Purpose.Radio.Television, Purpose.DomesticAppliance,Purpose.Repairs,Purpose.Education, Purpose.Retraining,Purpose.Business,SavingsAccountBonds.l SavingsAccountBonds.100.to.500,SavingsAccountBonds.500.to.1000, EmploymentDuration.gt.7,Personal.Male.Single,OtherDebt Property.RealEstate,Property.Insurance,Property.CarOther, Job.UnemployedUnskilled,Job.UnskilledResident,Job.SkilledEmplo

Subset the data

```
data <-  dplyr::select(GermanCredit,
    Amount,
    Duration,InstallmentRatePercentage,
    Telephone,Class,CheckingAccountStatus.lt.0,CheckingAccountStatus.gt.200,
    CreditHistory.NoCredit.AllPaid,Purpose.NewCar,Purpose.UsedCar,
    Purpose.Furniture.Equipment,Purpose.Radio.Television,
    Purpose.DomesticAppliance,Purpose.Repairs,Purpose.Education,
    Purpose.Retraining,Purpose.Business,SavingsAccountBonds.lt.100,
    SavingsAccountBonds.100.to.500,SavingsAccountBonds.500.to.1000,
    EmploymentDuration.gt.7,Personal.Male.Single,OtherDebtorsGuarantors.CoApplicant,
    Property.RealEstate,Property.Insurance,Property.CarOther,
    Job.UnemployedUnskilled,Job.UnskilledResident,Job.SkilledEmployee
  )
```

1. Split sample randomly into training-test using a 632:368 ratio.

```
# shuffle the data first
data <- data[sample(nrow(data)),]
train <- sample(seq_len(nrow(data)), size = floor(0.632*nrow(data)))
# split the data set into train and test
train_data <- data[train,]
test_data <- data[-train,]
```

2. Build the model using the 63.2% training data and compute R-square in holdout data. (function lm() gives R-squares)

```
fit <- lm(Amount~., data = train_data)
predicted <- predict(fit, test_data)
SSE <- sum((test_data$Amount- predicted) ^ 2)
SST <- sum((test_data$Amount - mean(test_data$Amount)) ^ 2)
r_sq_test <- 1 - (SSE/SST)
r_sq_test
```

```
## [1] 0.6269036
```

3. Save the coefficients, R-square in training and holdout samples. (To compute R-square in Holdout, take the square of correlation between actual and predicted values)

```
# coefficients
fit$coefficients
```

```
##                          (Intercept)                             Duration
##                            6897.6560                             118.9988
##           InstallmentRatePercentage                            Telephone
##                             -889.3917                             -573.8723
##                            ClassGood          CheckingAccountStatus.lt.0
##                             -471.3774                             -266.0254
##       CheckingAccountStatus.gt.200       CreditHistory.NoCredit.AllPaid
##                             -763.9888                             1081.5780
##                       Purpose.NewCar                     Purpose.UsedCar
##                             -732.1035                             -195.8142
##         Purpose.Furniture.Equipment           Purpose.Radio.Television
##                             -898.4256                            -1104.4497
##           Purpose.DomesticAppliance                       Purpose.Repairs
##                            -1514.0621                             -979.3084
##                    Purpose.Education                  Purpose.Retraining
##                            -1116.2877                             -793.5992
##                     Purpose.Business          SavingsAccountBonds.lt.100
##                             -948.1490                             -259.0806
##      SavingsAccountBonds.100.to.500     SavingsAccountBonds.500.to.1000
##                             -451.1011                             -597.3057
##             EmploymentDuration.gt.7               Personal.Male.Single
##                             -253.0326                              572.7284
## OtherDebtorsGuarantors.CoApplicant                 Property.RealEstate
##                              760.0712                            -1012.8500
##                  Property.Insurance                   Property.CarOther
##                             -697.6885                             -582.8182
##             Job.UnemployedUnskilled               Job.UnskilledResident
##                            -1487.7174                            -1394.3154
##                 Job.SkilledEmployee
##                            -1395.4937
```

```
hold_out_r2 <- cor(predicted, test_data$Amount)^2
hold_out_r2
```

```
## [1] 0.6358496
```

- **Repeat steps 1-3 1000 times. Save all 1000 results.**

```
Result <- data.frame(
  matrix(ncol = 32, nrow = 1000)
)
colnames(Result) <- c(
  "Intercept",
  colnames(data[,-1]),
  "R_Train",
  "R_Test",
  "Percent_r_fall"

)

for(i in 1:1000){

train <- sample(seq_len(nrow(data)), size = floor(0.632*nrow(data)))
# split the data set into train and test
```

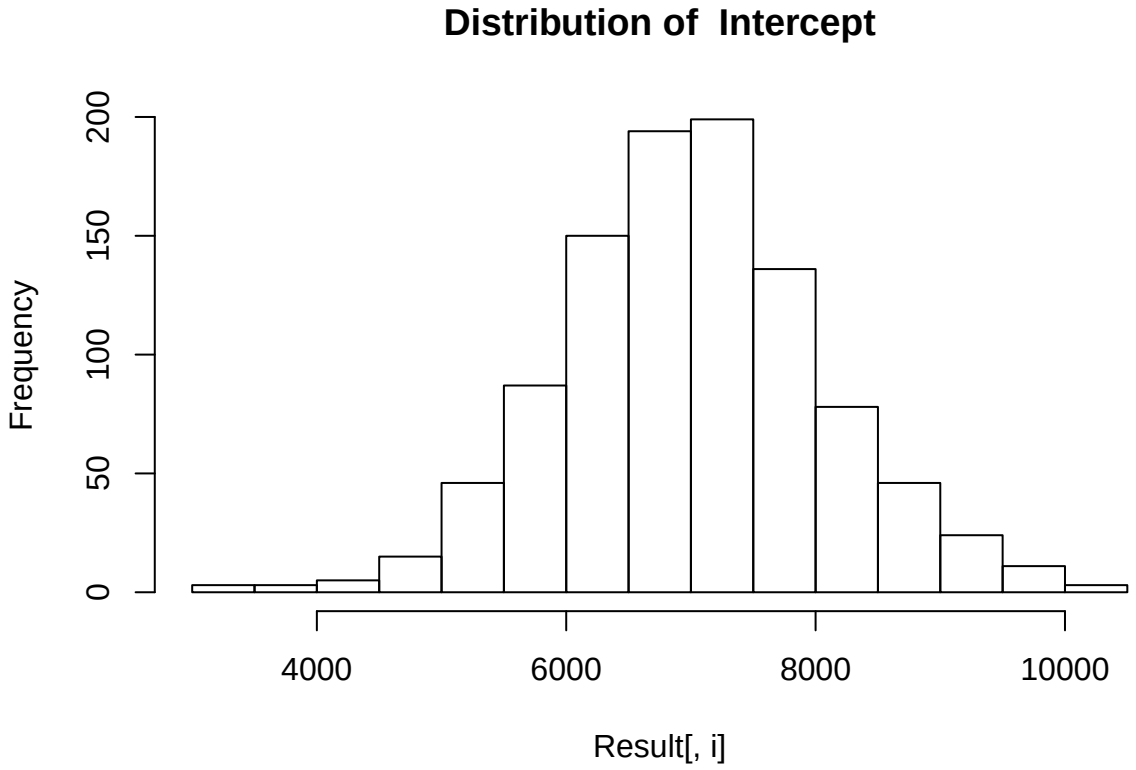
```
train_data <- data[train,]
test_data <- data[-train,]
fit <- lm(Amount~., data = train_data)
coefficients <- fit$coefficients
predicted <- predict(fit, test_data)
train_r2 <- summary(fit)$r.squared
# coefficients
fit$coefficients
hold_out_r2 <- cor(predicted, test_data$Amount)^2
hold_out_r2

R_fall <- (train_r2 - hold_out_r2)/train_r2
Result[i,] <-  c(coefficients, train_r2, hold_out_r2, R_fall)

}
head(Result)
```

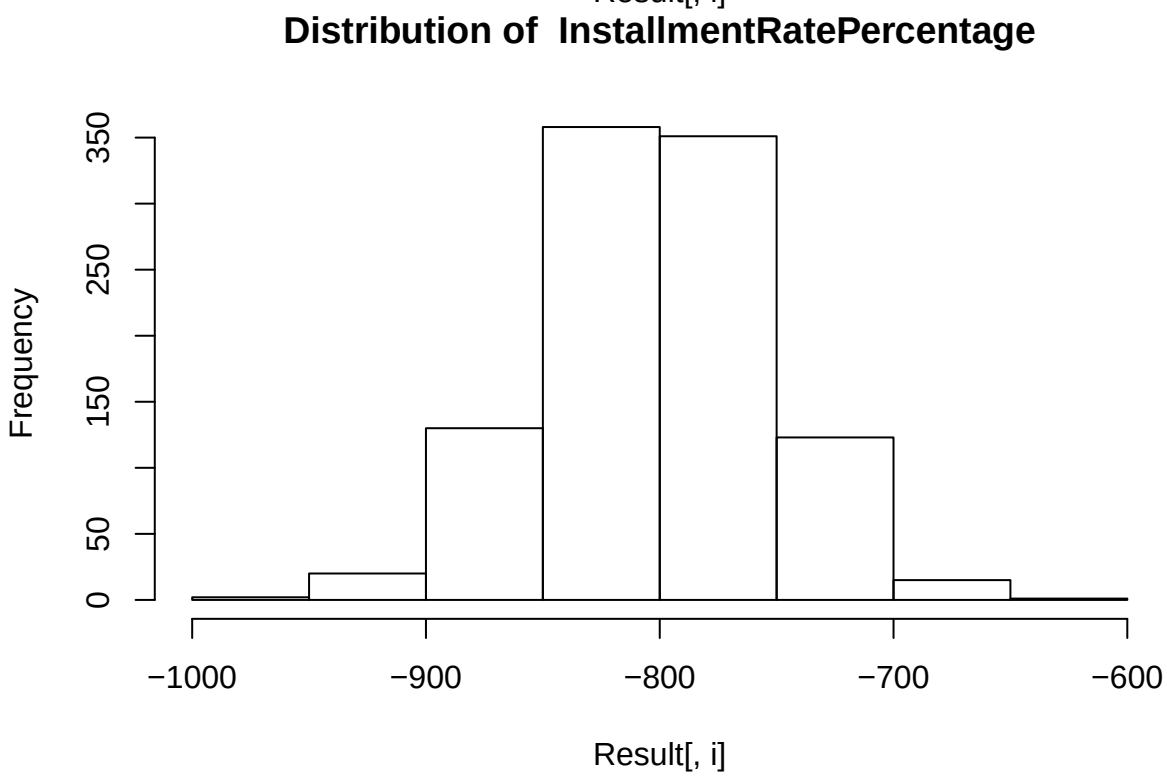
```
##   Intercept Duration InstallmentRatePercentage Telephone     Class
## 1  5997.402 125.0952                -814.4248 -558.3511 -266.8283
## 2  6684.430 121.8085                -730.5743 -595.6909 -625.2221
## 3  5837.186 132.4694                -847.0147 -369.0337 -395.4576
## 4  6500.453 120.9158                -798.9548 -462.0961 -398.3772
## 5  7000.015 121.3547                -806.1853 -388.8823 -469.6697
## 6  7313.078 121.6553                -778.6768 -444.8214 -452.7460
##   CheckingAccountStatus.lt.0 CheckingAccountStatus.gt.200
## 1                  -317.8220                   -744.5286
## 2                  -439.1568                   -509.7869
## 3                  -149.7021                   -604.6716
## 4                  -293.0251                   -907.9235
## 5                  -422.3590                   -772.8957
## 6                  -377.2467                   -621.2732
##   CreditHistory.NoCredit.AllPaid Purpose.NewCar Purpose.UsedCar
## 1                      1080.6926     -1179.2093      -713.95532
## 2                       999.2345      -993.3005       -63.49778
## 3                       320.9432     -1044.9790      -465.42405
## 4                      1025.0090      -622.1309      -112.06369
## 5                       193.3689     -1269.7511      -607.77201
## 6                       476.5450     -1595.4417     -1295.96176
##   Purpose.Furniture.Equipment Purpose.Radio.Television
## 1                  -1204.3682               -1240.3790
## 2                  -1158.5947               -1313.5791
## 3                  -1262.1951               -1398.4247
## 4                   -633.7962                -940.1085
## 5                  -1427.2019               -1591.7981
## 6                  -1667.4439               -2005.7060
##   Purpose.DomesticAppliance Purpose.Repairs Purpose.Education
## 1                -1806.4101       -902.5017        -1066.0182
## 2                 -988.3797      -1125.3196        -1330.0479
## 3                -1272.4309      -1001.1448        -1770.7121
## 4                -1479.6196      -1049.4595         -748.3886
## 5                -2556.3269      -1181.6328        -2174.6329
## 6                -2406.1688      -1814.4035        -1923.1753
##   Purpose.Retraining Purpose.Business SavingsAccountBonds.lt.100
## 1          -819.0465        -1190.295                  -359.9250
```

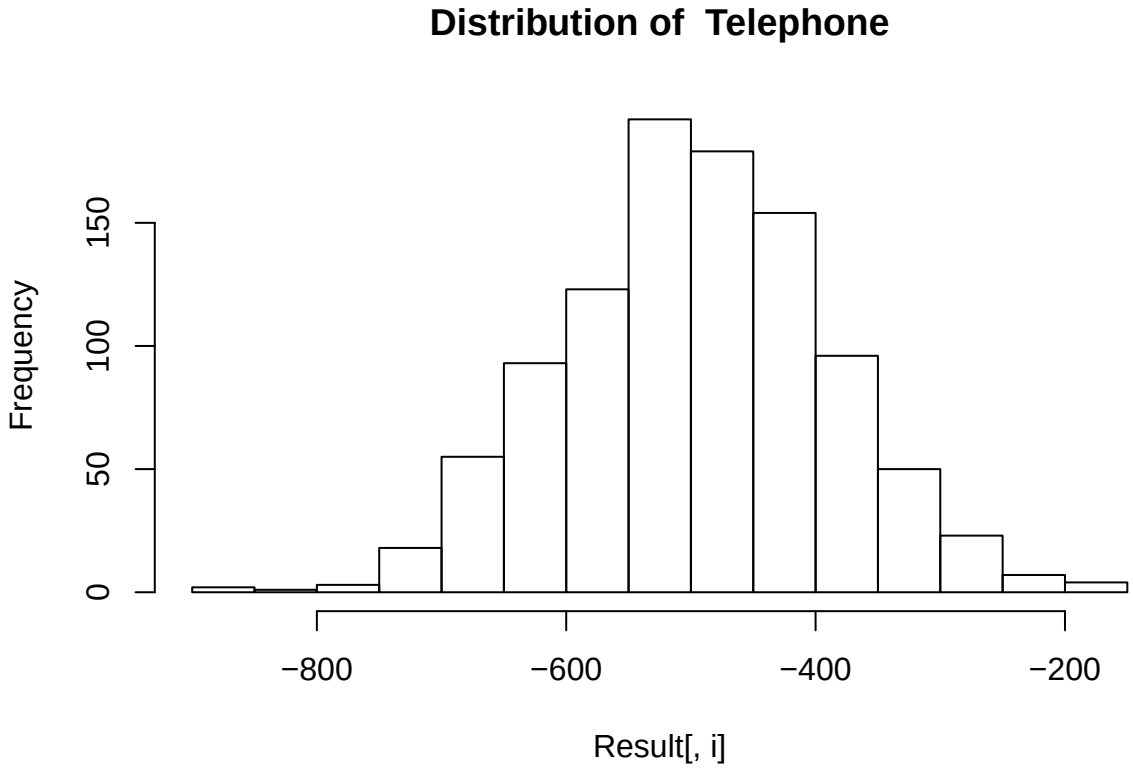
```
## 2         -2403.1181        -1134.174                  -179.7373
## 3         -1388.6376        -1107.524                  -187.2760
## 4          -358.9308         -578.719                  -402.3620
## 5         -1706.4712        -1317.308                  -295.2303
## 6         -2799.0021        -1592.974                  -362.6553
##   SavingsAccountBonds.100.to.500 SavingsAccountBonds.500.to.1000
## 1                      -510.6604                       -712.0310
## 2                      -502.0123                       -507.9058
## 3                      -593.8722                       -607.5310
## 4                      -747.3959                       -605.3814
## 5                      -985.0345                       -570.3024
## 6                      -804.8887                       -477.2529
##   EmploymentDuration.gt.7 Personal.Male.Single
## 1               -158.82695             426.9625
## 2               -288.64763             420.0193
## 3                -48.27112             615.6802
## 4                -71.44129             461.6524
## 5               -245.57082             542.6936
## 6               -371.10284             517.9234
##   OtherDebtorsGuarantors.CoApplicant Property.RealEstate
## 1                           445.8095           -691.4002
## 2                           992.5311           -921.1690
## 3                           364.8889           -345.1900
## 4                          1177.6141           -835.3026
## 5                           780.4325           -733.9850
## 6                           756.3335           -728.2684
##   Property.Insurance Property.CarOther Job.UnemployedUnskilled
## 1          -407.7869        -433.37633                -922.209
## 2          -665.7209        -731.02971               -2219.810
## 3          -101.0764          81.60426               -2587.422
## 4          -650.5651        -627.64144               -2128.816
## 5          -337.2238        -195.27177               -1882.372
## 6          -518.4063        -534.11536               -1713.712
##   Job.UnskilledResident Job.SkilledEmployee   R_Train    R_Test
## 1             -750.2321           -887.0702 0.6335544 0.5508576
## 2            -1220.8124          -1375.5174 0.6078334 0.5726184
## 3            -1280.2900          -1296.3271 0.6152682 0.5557454
## 4            -1694.2221          -1603.0839 0.6045863 0.5836248
## 5            -1618.0515          -1549.6315 0.5916877 0.5987551
## 6            -1507.3022          -1541.7658 0.5890638 0.6164582
##   Percent_r_fall
## 1     0.13052833
## 2     0.05793530
## 3     0.09674286
## 4     0.03467075
## 5    -0.01194452
## 6    -0.04650498
```

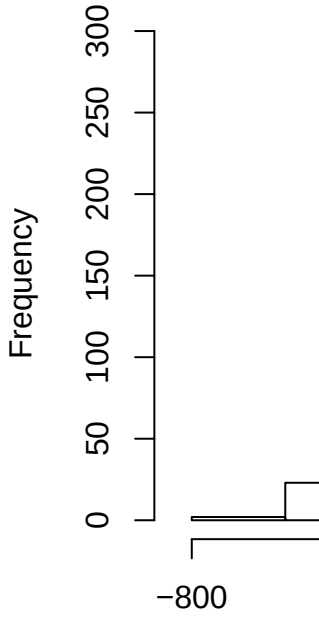
- Plot the distributions of all coefficients, holdout R2, and % fall in R2.

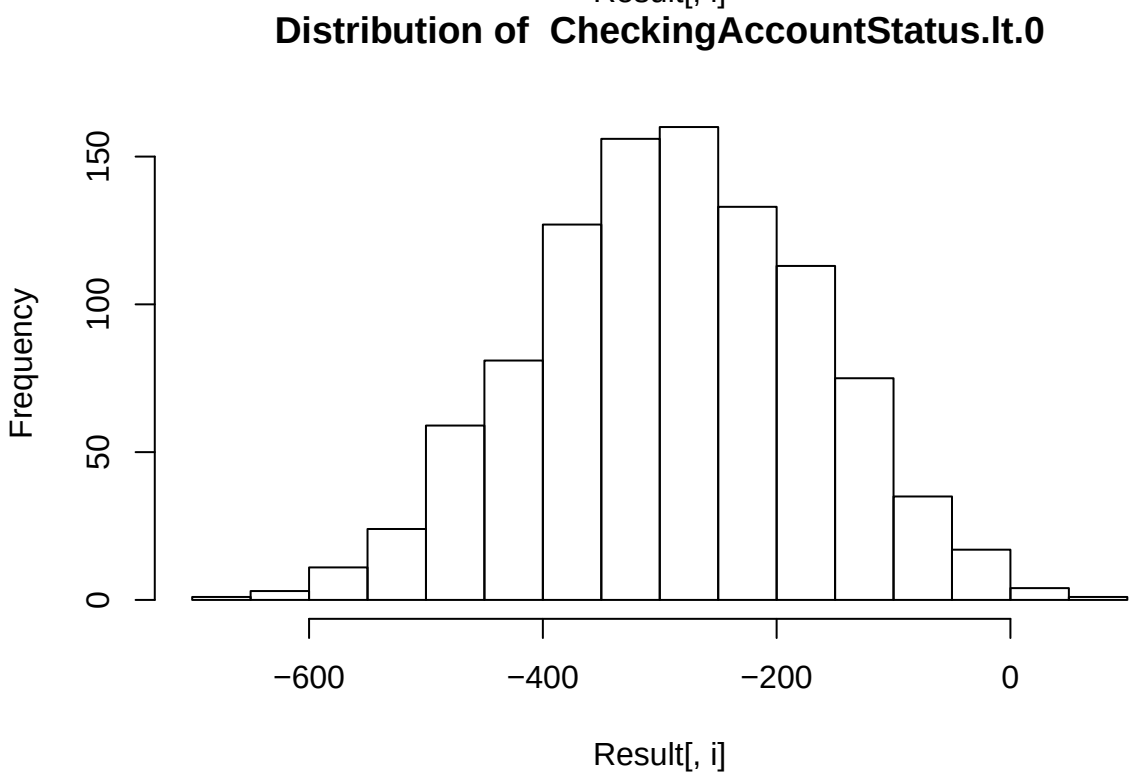
```
for(i in colnames(Result)){
  hist(Result[,i], main = paste("Distribution of ",i))
}
```

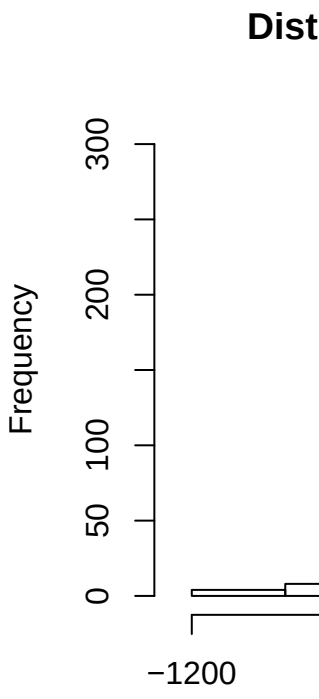

Distribution of Intercept
Frequency
0
50
100
150
200
4000
6000
8000
10000
Result[, i]

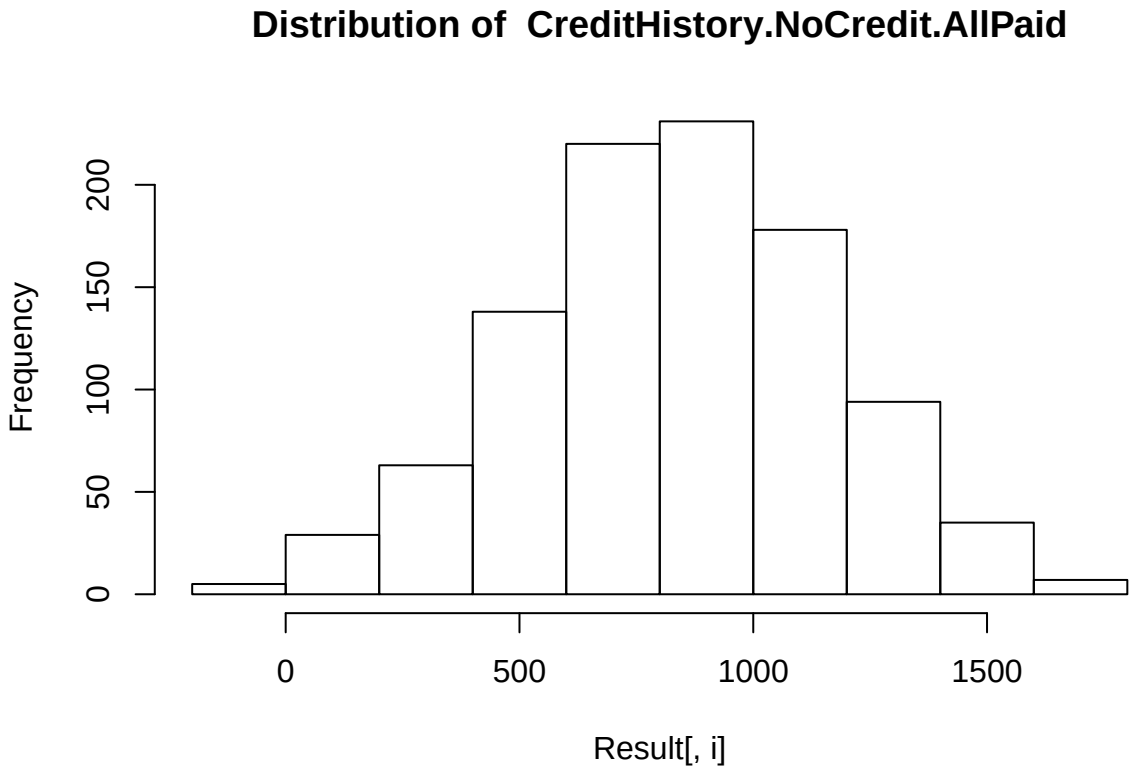

Distribution of InstallmentRatePercentage
Frequency
0
50
150
250
350
−1000
−900
−800
−700
−600
Result[, i]

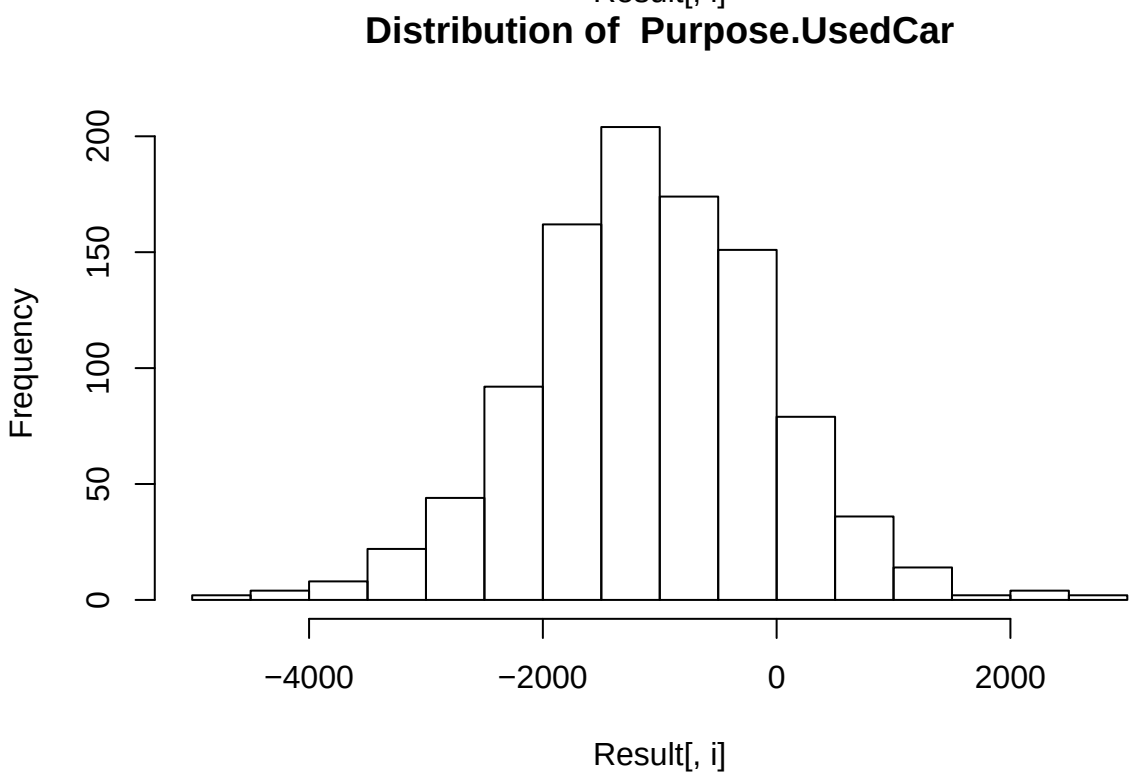

Distribution of Telephone
Frequency
0
50
100
150
−800
−600
−400
−200
Result[, i]

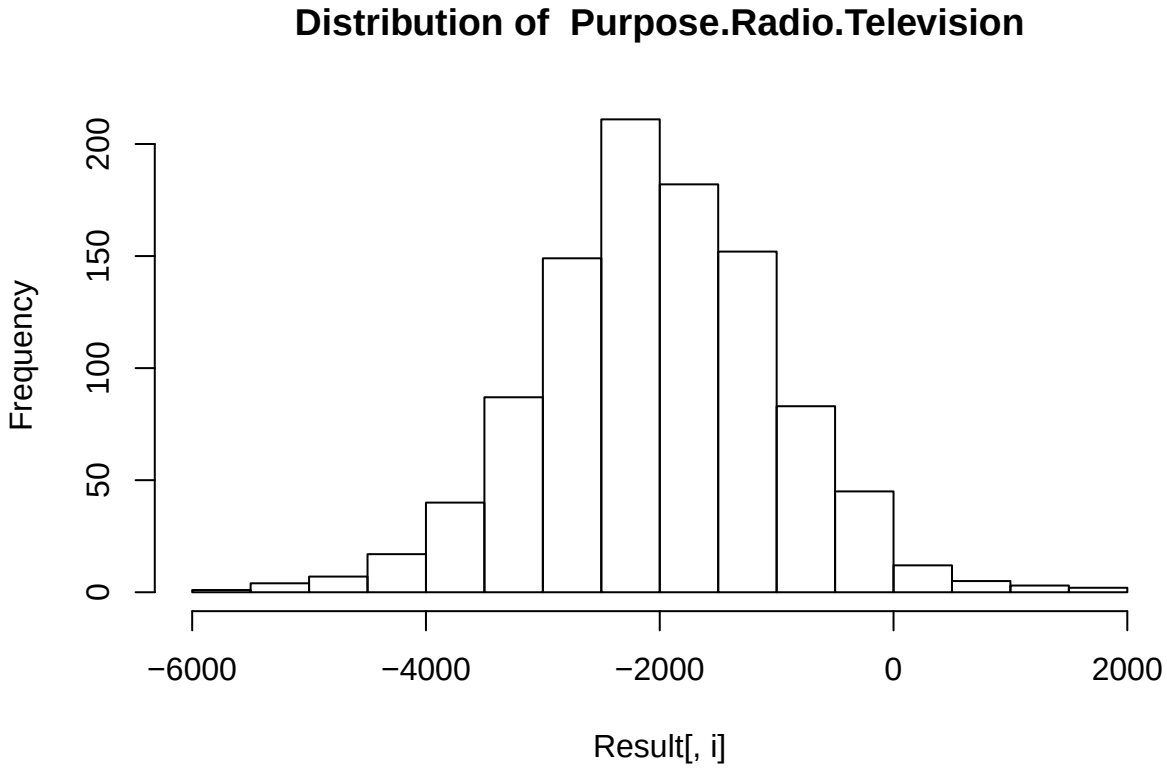

Frequency
0
50
100
150
200
250
300
−800

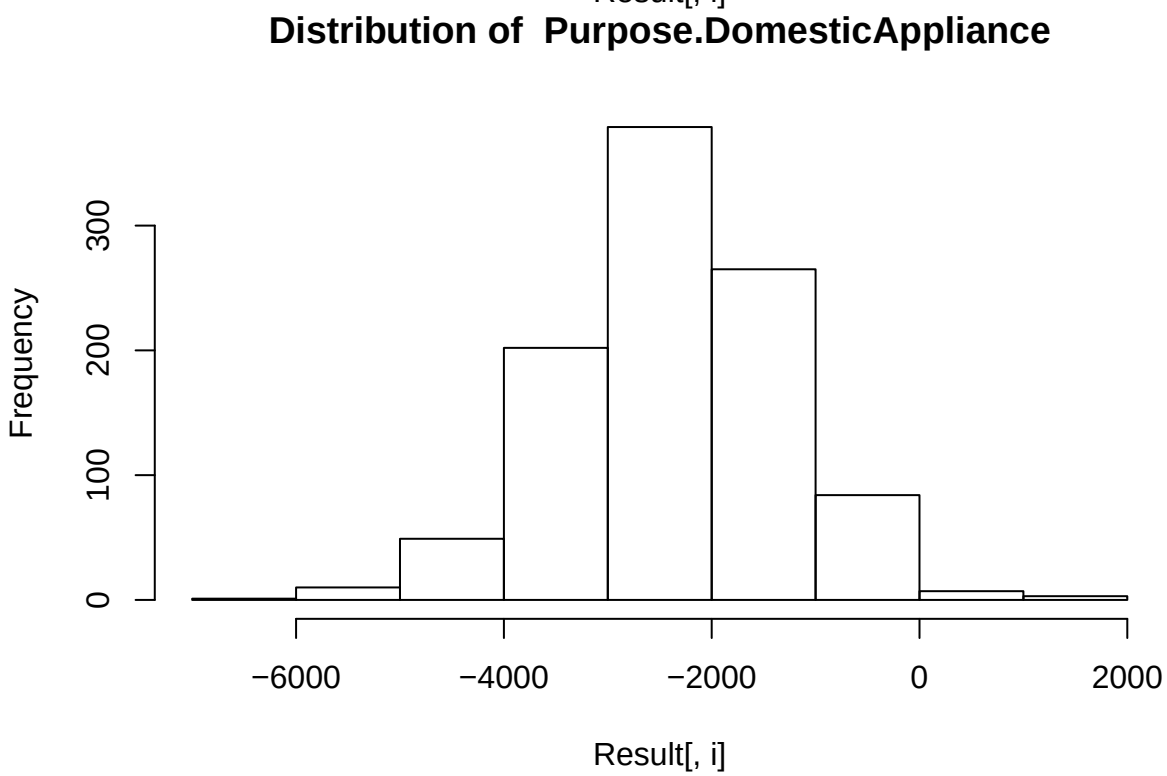

Distribution of CheckingAccountStatus.lt.0
Frequency
0
50
100
150
−600
−400
−200
0
Result[, i]

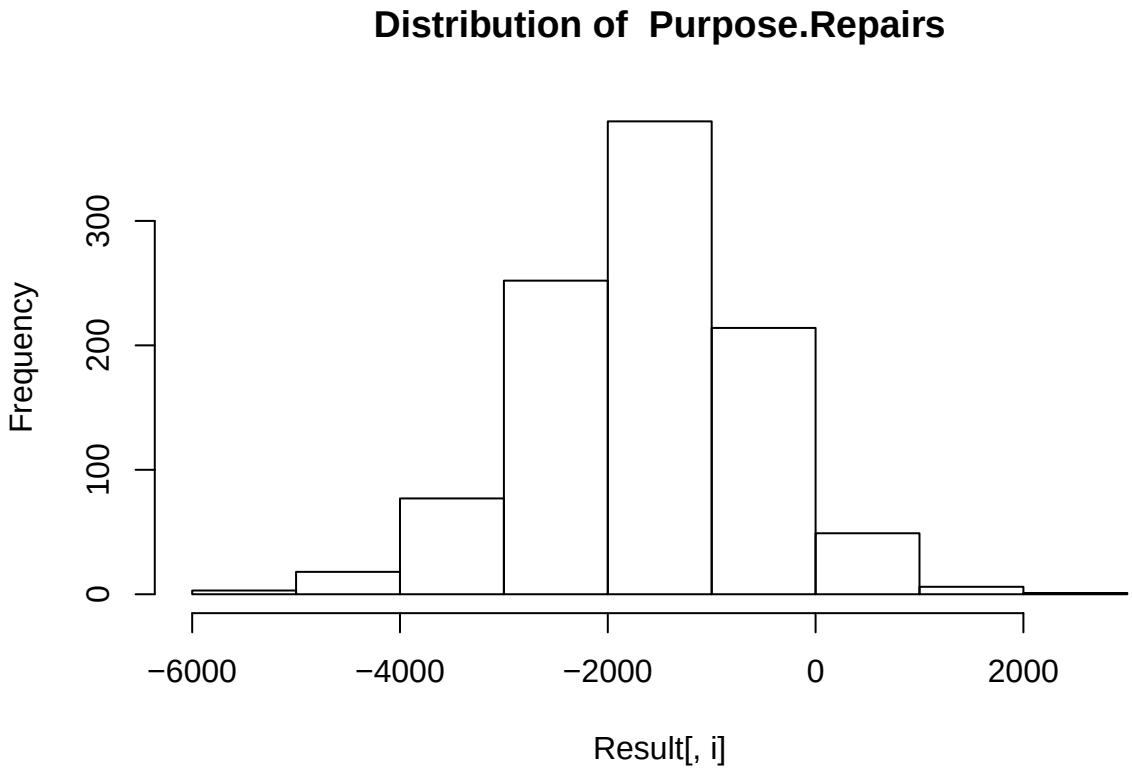

Dist
Frequency
0
50
100
200
300
−1200

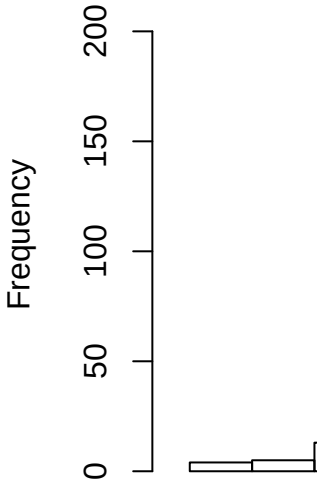

Distribution of CreditHistory.NoCredit.AllPaid
Frequency
0
50
100
150
200
0
500
1000
1500
Result[, i]

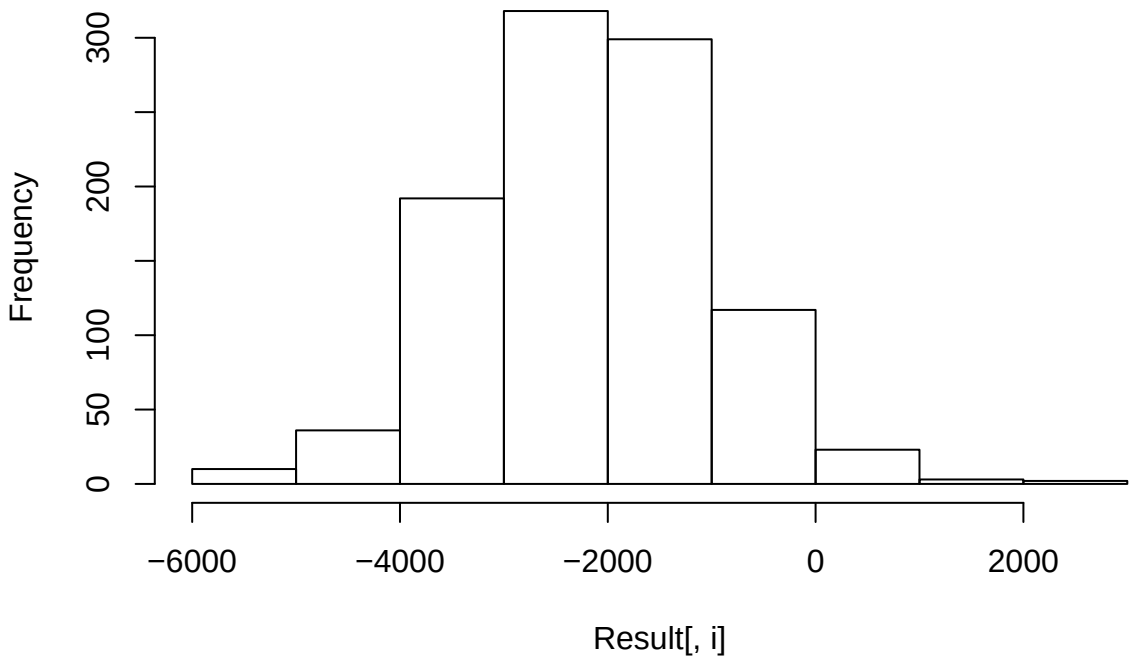

Distribution of Purpose.UsedCar
Frequency
0
50
100
150
200
−4000
−2000
0
2000
Result[, i]

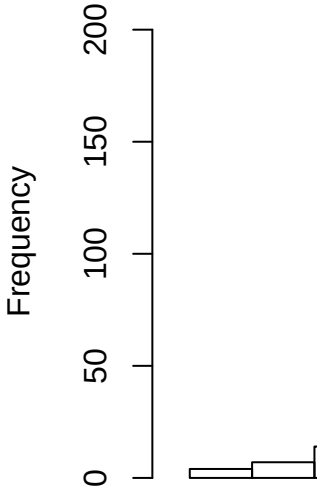

Distribution of Purpose.Radio.Television
Frequency
0
50
100
150
200
−6000
−4000
−2000
0
2000
Result[, i]

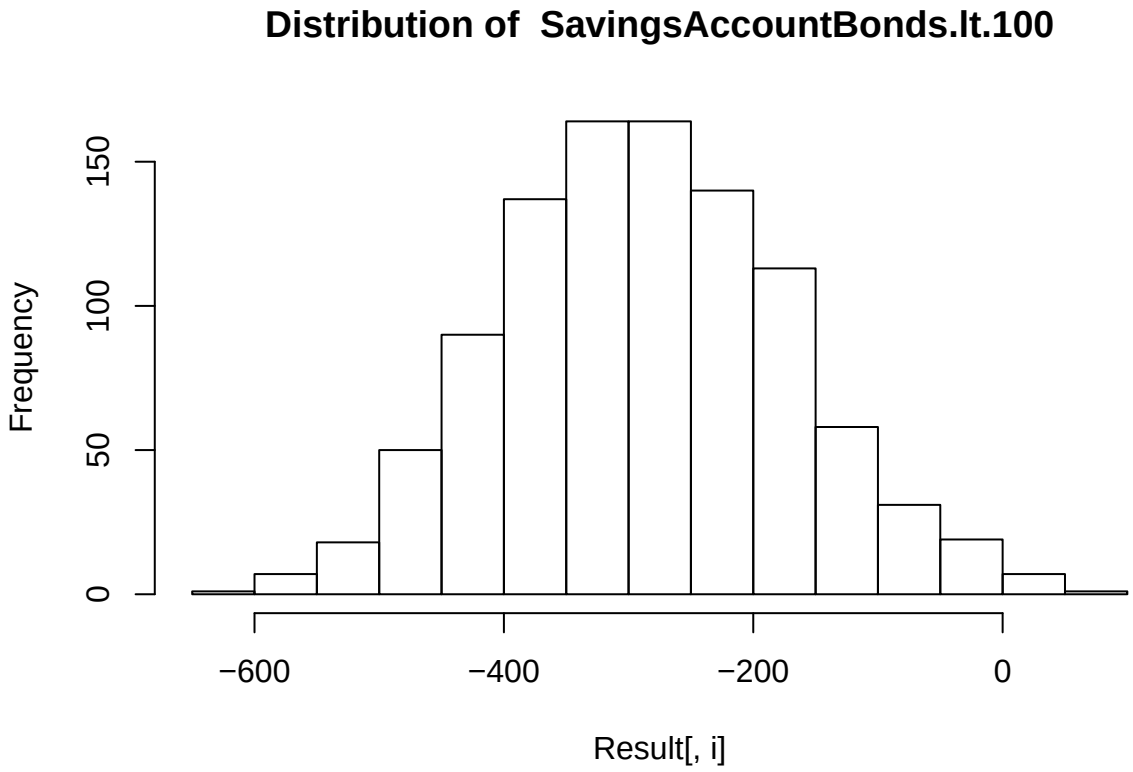

Distribution of Purpose.DomesticAppliance
Frequency
0
100
200
300
−6000
−4000
−2000
0
2000
Result[, i]

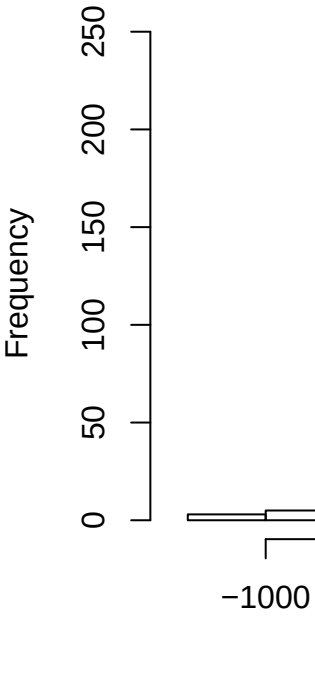

Distribution of Purpose.Repairs
Frequency
0
100
200
300
−6000
−4000
−2000
0
2000
Result[, i]

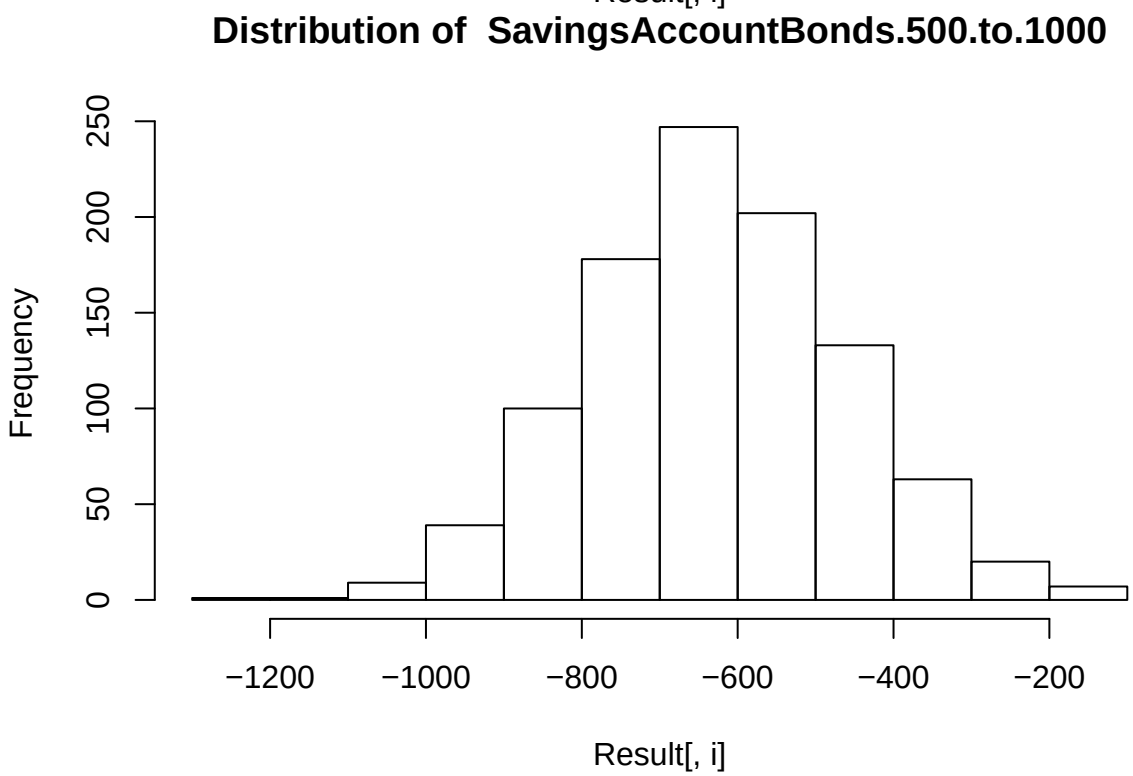

Frequency
0
50
100
150
200

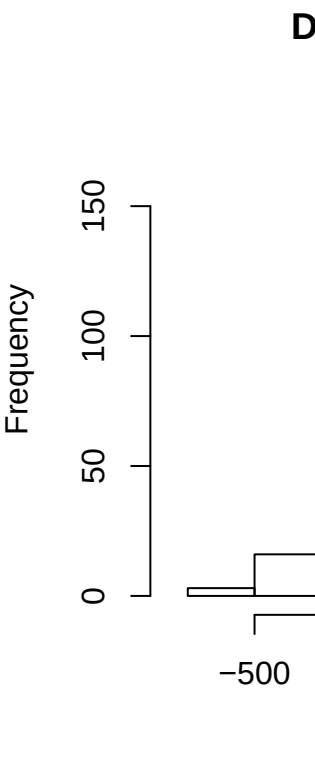

Distribution of Purpose.Retraining
Frequency
0
50
100
200
300
−6000
−4000
−2000
0
2000
Result[, i]

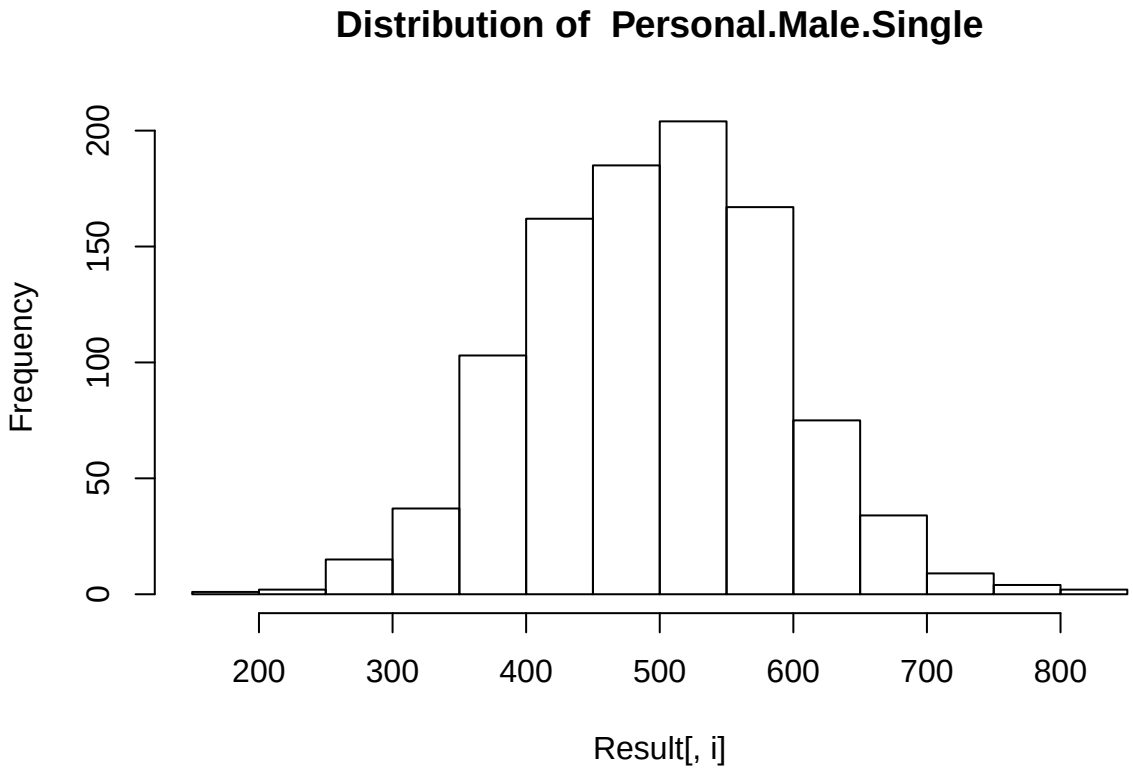

Frequency
0
50
100
150
200

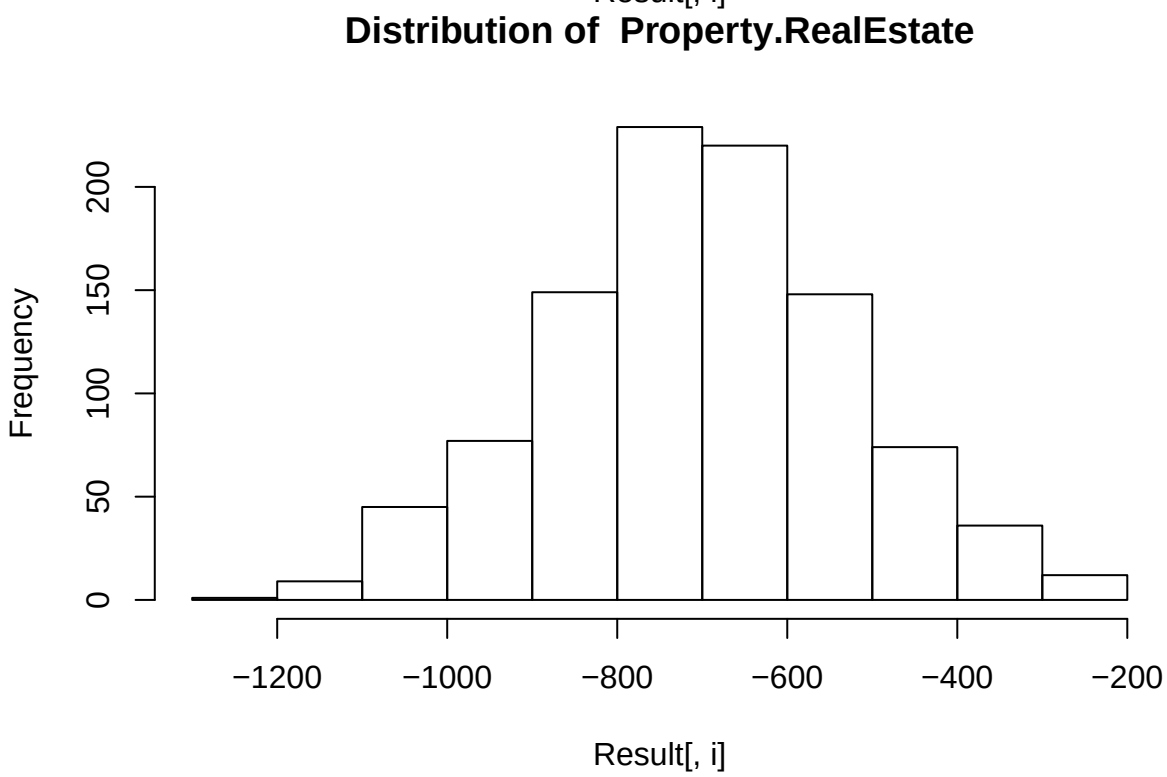

Distribution of SavingsAccountBonds.lt.100
Frequency
0
50
100
150
−600
−400
−200
0
Result[, i]

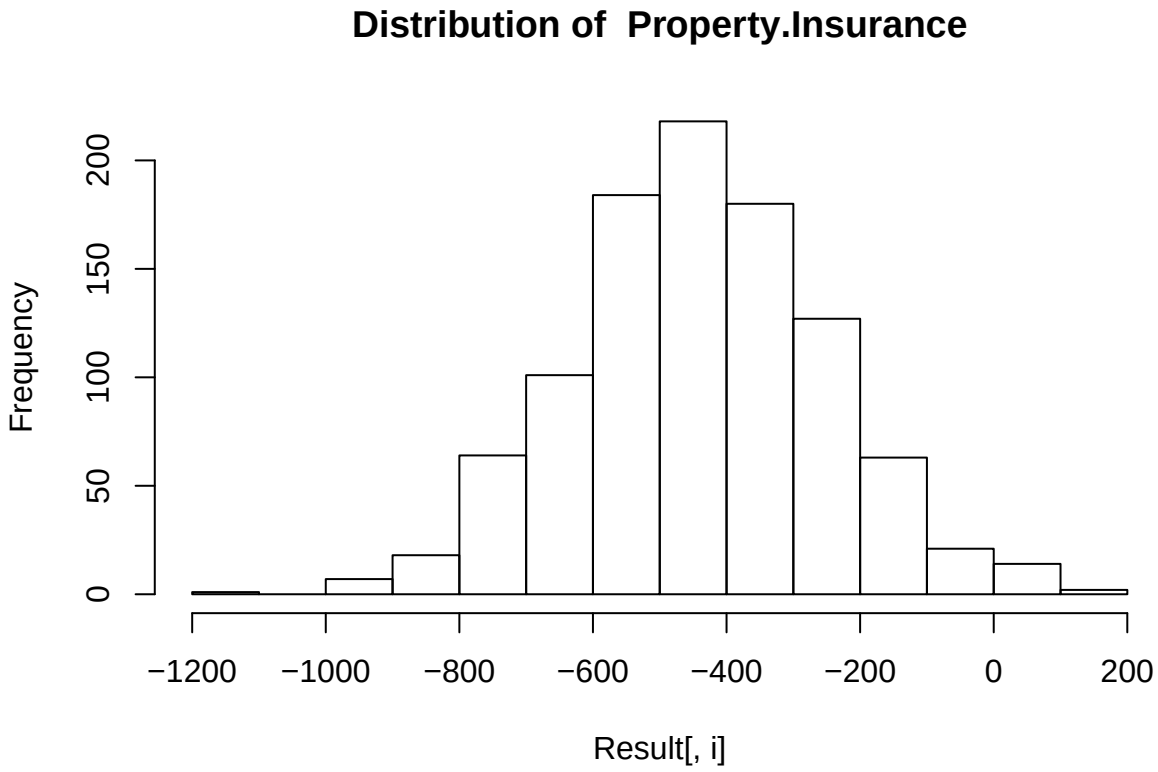

Distri
Frequency
0
50
100
150
200
250
−1000

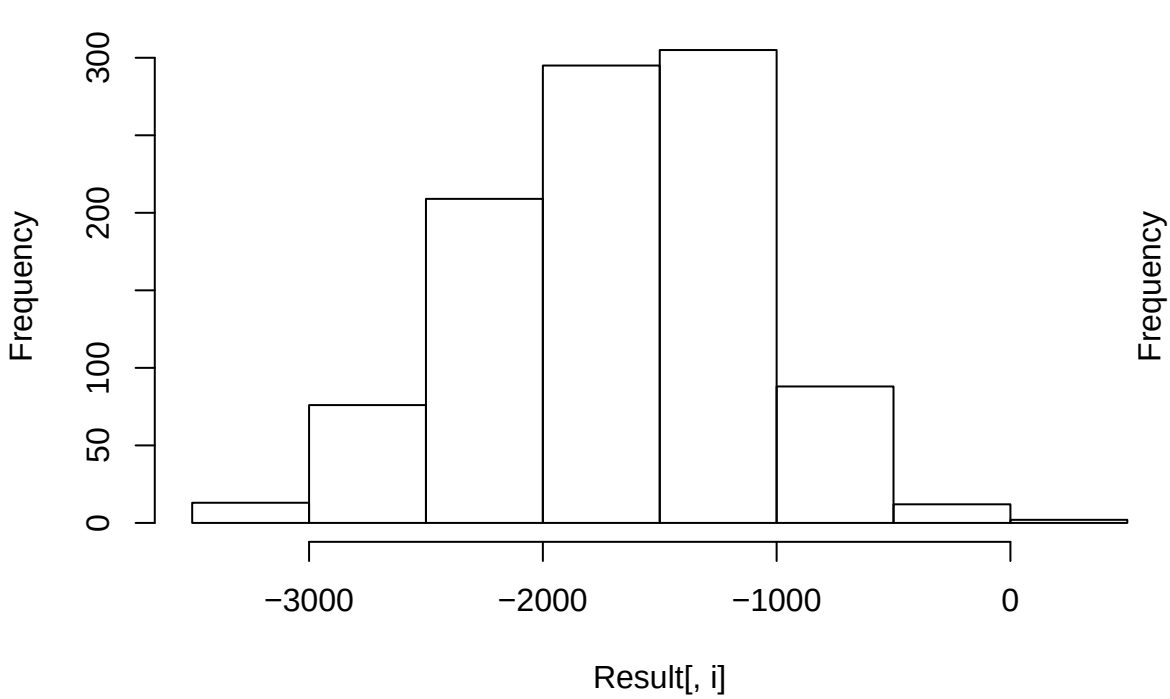

Distribution of SavingsAccountBonds.500.to.1000
Frequency
0
50
100
150
200
250
−1200
−1000
−800
−600
−400
−200
Result[, i]

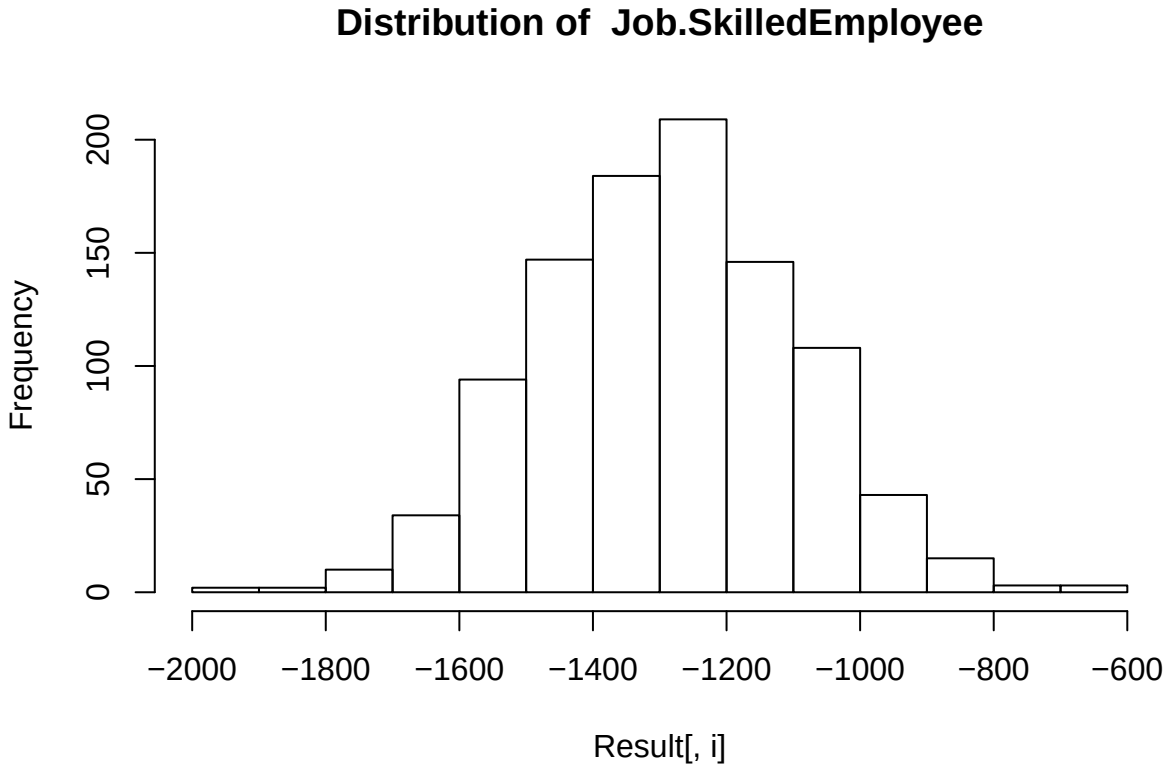

D
Frequency
0
50
100
150
−500

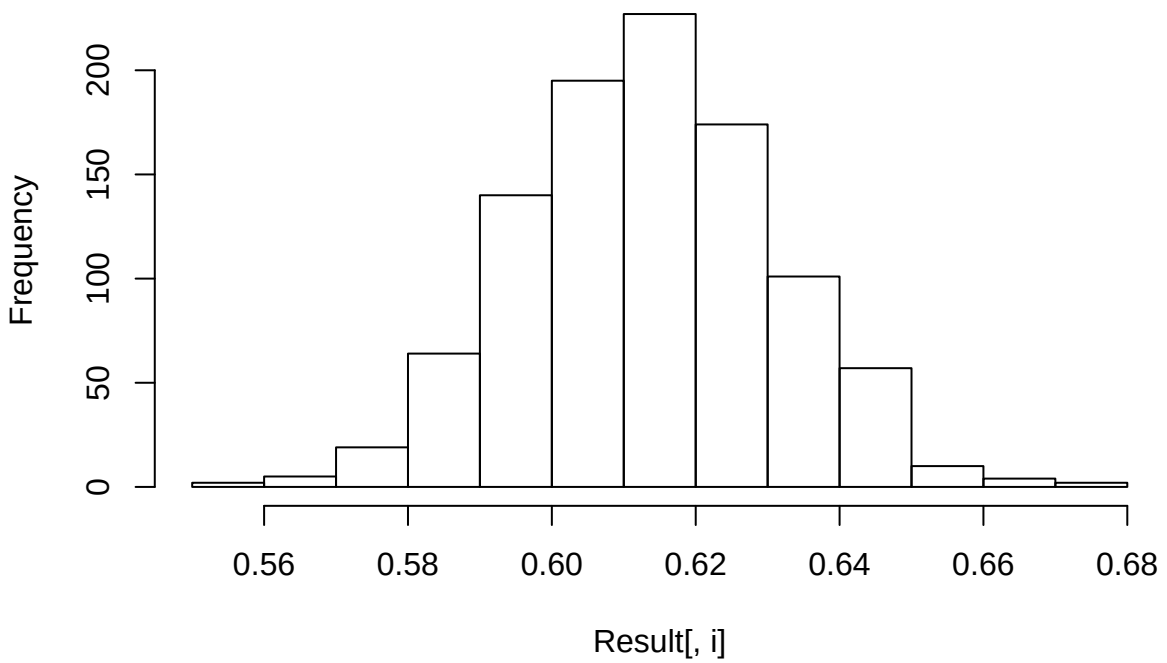

Distribution of Personal.Male.Single
Frequency
0
50
100
150
200
200
300
400
500
600
700
800
Result[, i]

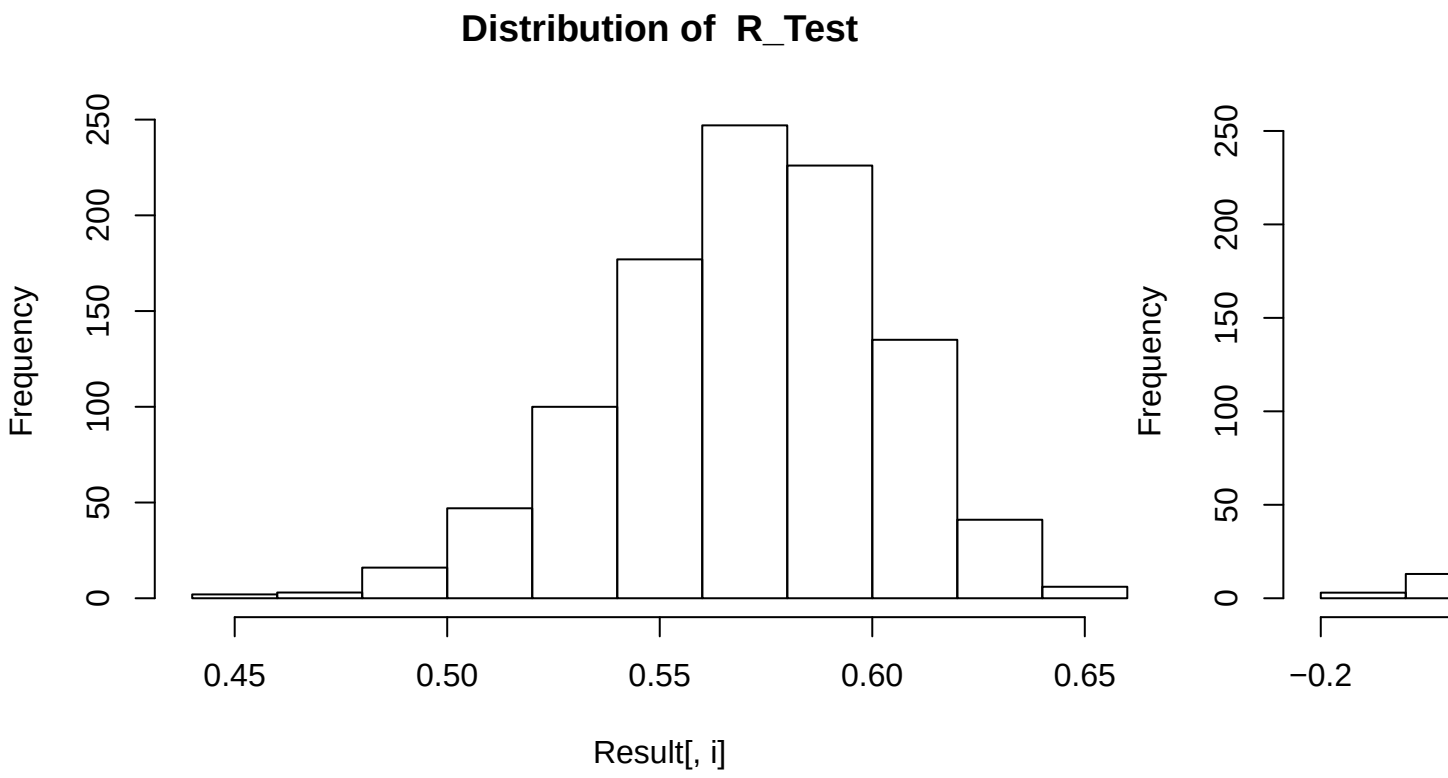

Distribution of Property.RealEstate
Frequency
0
50
100
150
200
−1200
−1000
−800
−600
−400
−200
Result[, i]


Distribution of Property.Insurance
Frequency
0
50
100
150
200
−1200
−1000
−800
−600
−400
−200
0
200
Result[, i]


Distribution of Job.UnemployedUnskilled
Frequency
0
50
100
200
300
−3000
−2000
−1000
0
Result[, i]


Distribution of Job.SkilledEmployee
Frequency
0
50
100
150
200
−2000
−1800
−1600
−1400
−1200
−1000
−800
−600
Result[, i]


Distribution of R_Train
Frequency
0
50
100
150
200
0.56
0.58
0.60
0.62
0.64
0.66
0.68
Result[, i]



- **Compute the averages of all 1000 coefficients.**

```
df.avg.sd <- data.frame(Term = colnames(Result[1:29]),Average = colMeans(Result[,1:29]))
rownames(df.avg.sd)<- NULL
df.avg.sd
```

```
##                                  Term    Average
## 1                           Intercept  6999.7707
## 2                            Duration   125.2071
## 3          InstallmentRatePercentage  -802.1857
## 4                           Telephone  -496.5434
## 5                               Class  -351.9009
## 6         CheckingAccountStatus.lt.0  -287.8906
## 7       CheckingAccountStatus.gt.200  -677.4092
## 8     CreditHistory.NoCredit.AllPaid   836.8170
## 9                     Purpose.NewCar -1724.9006
## 10                   Purpose.UsedCar -1074.9051
## 11       Purpose.Furniture.Equipment -1771.1696
## 12          Purpose.Radio.Television -2014.6744
## 13        Purpose.DomesticAppliance -2363.8638
## 14                   Purpose.Repairs -1645.7298
## 15                 Purpose.Education -1874.0647
## 16               Purpose.Retraining -2181.6235
## 17                 Purpose.Business -1911.7808
## 18       SavingsAccountBonds.lt.100  -287.2195
## 19    SavingsAccountBonds.100.to.500  -524.6950
## 20   SavingsAccountBonds.500.to.1000  -627.1409
## 21           EmploymentDuration.gt.7  -219.6476
```

```
## 22               Personal.Male.Single   496.1503
## 23 OtherDebtorsGuarantors.CoApplicant   619.7911
## 24                Property.RealEstate  -702.6036
## 25                 Property.Insurance  -438.9903
## 26                  Property.CarOther  -439.3459
## 27           Job.UnemployedUnskilled -1682.5699
## 28             Job.UnskilledResident -1241.1980
## 29              Job.SkilledEmployee -1288.5289
```

- **Compute the standard deviation of all 1000 coefficients (for each beta)**

```
sd <- sapply(Result[,1:29], sd)
df.avg.sd$Standard.Deviation<-sd
rownames(df.avg.sd)<- NULL
df.avg.sd
```

```
##                                Term    Average Standard.Deviation
## 1                         Intercept  6999.7707        1059.039936
## 2                          Duration   125.2071           5.615298
## 3       InstallmentRatePercentage  -802.1857          48.012199
## 4                         Telephone  -496.5434         106.846802
## 5                             Class  -351.9009         132.156832
## 6       CheckingAccountStatus.lt.0  -287.8906         120.914641
## 7     CheckingAccountStatus.gt.200  -677.4092         133.331920
## 8   CreditHistory.NoCredit.AllPaid   836.8170         328.356334
## 9                    Purpose.NewCar -1724.9006        1023.083635
## 10                  Purpose.UsedCar -1074.9051        1043.623469
## 11     Purpose.Furniture.Equipment -1771.1696        1031.437530
## 12       Purpose.Radio.Television -2014.6744        1029.888690
## 13      Purpose.DomesticAppliance -2363.8638        1055.886078
## 14                  Purpose.Repairs -1645.7298        1079.039626
## 15                Purpose.Education -1874.0647        1032.600553
## 16               Purpose.Retraining -2181.6235        1148.174321
## 17                 Purpose.Business -1911.7808        1035.405441
## 18      SavingsAccountBonds.lt.100  -287.2195         115.226941
## 19   SavingsAccountBonds.100.to.500  -524.6950         161.836457
## 20  SavingsAccountBonds.500.to.1000  -627.1409         166.965630
## 21          EmploymentDuration.gt.7  -219.6476         104.969506
## 22             Personal.Male.Single   496.1503          93.175928
## 23 OtherDebtorsGuarantors.CoApplicant   619.7911         243.019630
## 24              Property.RealEstate  -702.6036         177.589072
## 25               Property.Insurance  -438.9903         189.180728
## 26                Property.CarOther  -439.3459         184.388860
## 27          Job.UnemployedUnskilled -1682.5699         582.962092
## 28            Job.UnskilledResident -1241.1980         212.385629
## 29             Job.SkilledEmployee -1288.5289         192.833279
```

- **Compare average across 1000 to single model built using entire sample.**

Build model on entire data

```r
fit <- lm(Amount~., data = data)
df.avg.sd$Full_model <- fit$coefficients
df.avg.sd
```

```
##                                 Term    Average Standard.Deviation
## 1                          Intercept  6999.7707        1059.039936
## 2                           Duration   125.2071           5.615298
## 3       InstallmentRatePercentage  -802.1857          48.012199
## 4                          Telephone  -496.5434         106.846802
## 5                              Class  -351.9009         132.156832
## 6       CheckingAccountStatus.lt.0  -287.8906         120.914641
## 7     CheckingAccountStatus.gt.200  -677.4092         133.331920
## 8   CreditHistory.NoCredit.AllPaid   836.8170         328.356334
## 9                     Purpose.NewCar -1724.9006        1023.083635
## 10                   Purpose.UsedCar -1074.9051        1043.623469
## 11      Purpose.Furniture.Equipment -1771.1696        1031.437530
## 12        Purpose.Radio.Television -2014.6744        1029.888690
## 13       Purpose.DomesticAppliance -2363.8638        1055.886078
## 14                   Purpose.Repairs -1645.7298        1079.039626
## 15                 Purpose.Education -1874.0647        1032.600553
## 16               Purpose.Retraining -2181.6235        1148.174321
## 17                 Purpose.Business -1911.7808        1035.405441
## 18      SavingsAccountBonds.lt.100  -287.2195         115.226941
## 19  SavingsAccountBonds.100.to.500  -524.6950         161.836457
## 20 SavingsAccountBonds.500.to.1000  -627.1409         166.965630
## 21          EmploymentDuration.gt.7  -219.6476         104.969506
## 22             Personal.Male.Single   496.1503          93.175928
## 23 OtherDebtorsGuarantors.CoApplicant   619.7911         243.019630
## 24               Property.RealEstate  -702.6036         177.589072
## 25                Property.Insurance  -438.9903         189.180728
## 26                Property.CarOther  -439.3459         184.388860
## 27          Job.UnemployedUnskilled -1682.5699         582.962092
## 28            Job.UnskilledResident -1241.1980         212.385629
## 29              Job.SkilledEmployee -1288.5289         192.833279
##    Full_model
## 1   7026.7696
## 2    125.2081
## 3   -802.7376
## 4   -499.8240
## 5   -353.3310
## 6   -285.1747
## 7   -672.2566
## 8    829.4857
## 9  -1741.6057
## 10 -1084.1573
## 11 -1783.7420
## 12 -2032.6309
## 13 -2388.6691
## 14 -1663.7689
## 15 -1886.9521
## 16 -2205.5276
## 17 -1926.9312
## 18  -287.4003
## 19  -517.5188
```

```
## 20  -623.1515
## 21  -218.6188
## 22   487.9634
## 23   618.3677
## 24  -700.9964
## 25  -444.7788
## 26  -444.1896
## 27 -1683.7167
## 28 -1245.1894
## 29 -1290.8688
```

```
mean(Result$R_Train)
```

```
## [1] 0.6135389
```

```
summary(fit)$r.squared
```

```
## [1] 0.6058165
```

Average across 1000 sample coefficients somewhat approximately close to single model built using entire sample. Train R square is close to the entire sample r square.

- **Sort each coefficient's 1000 values. Compute 2.5%-97.5% Confidence Intervals (CI). Scale these CI's down by a factor of .632^0.5 .**

Sort the data

```
sorted_result <- apply(Result,2, sort)
head(sorted_result)
```

```
##      Intercept Duration InstallmentRatePercentage Telephone     Class
## [1,]  3253.210 105.7410                -966.2513 -876.3390 -761.6926
## [2,]  3347.349 106.9432                -964.1438 -854.9945 -701.6113
## [3,]  3354.516 108.2757                -933.0598 -802.7194 -697.0504
## [4,]  3702.396 108.6229                -932.5395 -786.3900 -690.1568
## [5,]  3809.508 109.5847                -930.9516 -767.5740 -672.4078
## [6,]  3884.605 109.6106                -929.1248 -760.6492 -664.6506
##      CheckingAccountStatus.lt.0 CheckingAccountStatus.gt.200
## [1,]                  -675.1657                    -1147.028
## [2,]                  -605.6603                    -1137.033
## [3,]                  -605.5736                    -1130.921
## [4,]                  -605.5605                    -1117.647
## [5,]                  -585.3195                    -1084.954
## [6,]                  -584.5056                    -1043.059
##      CreditHistory.NoCredit.AllPaid Purpose.NewCar Purpose.UsedCar
## [1,]                    -160.548080      -5268.550       -4725.240
## [2,]                     -79.280674      -5250.021       -4542.924
## [3,]                     -43.793317      -5126.855       -4357.603
## [4,]                     -12.207176      -4928.336       -4079.979
## [5,]                      -9.961633      -4736.334       -4042.432
## [6,]                       1.315482      -4626.589       -4001.666
##      Purpose.Furniture.Equipment Purpose.Radio.Television
## [1,]                   -5216.914                -5517.348
```

```
## [2,]                  -5137.725               -5450.723
## [3,]                  -5122.926               -5421.674
## [4,]                  -4975.010               -5143.764
## [5,]                  -4766.008               -5085.689
## [6,]                  -4557.507               -4883.626
##      Purpose.DomesticAppliance Purpose.Repairs Purpose.Education
## [1,]                -6140.265       -5566.616         -5410.052
## [2,]                -5935.224       -5548.195         -5298.199
## [3,]                -5352.040       -5223.399         -5201.936
## [4,]                -5350.585       -4961.716         -5123.504
## [5,]                -5331.050       -4726.904         -4676.047
## [6,]                -5271.151       -4719.708         -4657.647
##      Purpose.Retraining Purpose.Business SavingsAccountBonds.lt.100
## [1,]          -5992.407        -5273.549                  -616.9432
## [2,]          -5729.106        -5239.763                  -585.5442
## [3,]          -5447.725        -5115.548                  -582.3733
## [4,]          -5380.247        -5035.589                  -570.2351
## [5,]          -5366.826        -4760.574                  -562.7742
## [6,]          -5307.625        -4739.781                  -559.6871
##      SavingsAccountBonds.100.to.500 SavingsAccountBonds.500.to.1000
## [1,]                     -1098.1993                        -1253.896
## [2,]                     -1046.9787                        -1141.383
## [3,]                     -1004.6244                        -1085.878
## [4,]                      -985.0345                        -1044.568
## [5,]                      -939.0509                        -1040.265
## [6,]                      -927.2905                        -1038.709
##      EmploymentDuration.gt.7 Personal.Male.Single
## [1,]               -525.1191             192.7818
## [2,]               -519.2421             222.6325
## [3,]               -513.0029             241.4897
## [4,]               -499.8206             260.5351
## [5,]               -498.5323             266.4648
## [6,]               -498.0054             267.8687
##      OtherDebtorsGuarantors.CoApplicant Property.RealEstate
## [1,]                          -220.21850           -1237.945
## [2,]                          -171.17045           -1198.257
## [3,]                           -97.53525           -1189.837
## [4,]                           -68.02736           -1165.636
## [5,]                           -17.08723           -1164.077
## [6,]                            45.17373           -1154.174
##      Property.Insurance Property.CarOther Job.UnemployedUnskilled
## [1,]         -1108.8362        -1024.7486                -3352.868
## [2,]          -959.4551         -961.1962                -3273.900
## [3,]          -959.3375         -921.5723                -3254.847
## [4,]          -938.6373         -910.8777                -3246.849
## [5,]          -913.1702         -909.0438                -3199.276
## [6,]          -910.5408         -907.2624                -3087.022
##      Job.UnskilledResident Job.SkilledEmployee   R_Train    R_Test
## [1,]             -2014.492           -1981.570 0.5508844 0.4510956
## [2,]             -1908.826           -1903.148 0.5554203 0.4556806
## [3,]             -1883.523           -1893.395 0.5669767 0.4627251
## [4,]             -1840.365           -1830.663 0.5677889 0.4739278
## [5,]             -1808.213           -1795.356 0.5684904 0.4760309
## [6,]             -1767.441           -1774.037 0.5686675 0.4823055
```

```
##      Percent_r_fall
## [1,]     -0.1882688
## [2,]     -0.1772130
## [3,]     -0.1526886
## [4,]     -0.1326346
## [5,]     -0.1253421
## [6,]     -0.1249709
```

Calculate 2.5%-97.5% confidence interval

```
lower_ci <- sapply(Result[,1:29], function(a){
  mean(a) - qnorm(0.975)*sd(a)/sqrt(1000)
})


upper_ci <- sapply(Result[,1:29], function(a){
  mean(a) + qnorm(0.975)*sd(a)/sqrt(1000)
})
print("2.5%-97.5% confidence interval")
```

```
## [1] "2.5%-97.5% confidence interval"
```

```
ci <- data.frame(Term = colnames(Result)[1:29],"lower 2.5%"= lower_ci,"upper 97.5%"= upper_ci)
rownames(ci)<- NULL
# scale the CI
ci
```

```
##                                 Term lower.2.5. upper.97.5.
## 1                          Intercept  6934.1319   7065.4094
## 2                           Duration   124.8591    125.5551
## 3          InstallmentRatePercentage  -805.1615   -799.2100
## 4                          Telephone  -503.1657   -489.9211
## 5                              Class  -360.0919   -343.7098
## 6        CheckingAccountStatus.lt.0  -295.3848   -280.3964
## 7       CheckingAccountStatus.gt.200  -685.6730   -669.1453
## 8     CreditHistory.NoCredit.AllPaid   816.4657    857.1684
## 9                     Purpose.NewCar -1788.3108  -1661.4904
## 10                   Purpose.UsedCar -1139.5883  -1010.2218
## 11       Purpose.Furniture.Equipment -1835.0976  -1707.2416
## 12          Purpose.Radio.Television -2078.5064  -1950.8424
## 13         Purpose.DomesticAppliance -2429.3071  -2298.4205
## 14                   Purpose.Repairs -1712.6081  -1578.8514
## 15                 Purpose.Education -1938.0647  -1810.0646
## 16               Purpose.Retraining -2252.7868  -2110.4602
## 17                 Purpose.Business -1975.9547  -1847.6069
## 18        SavingsAccountBonds.lt.100  -294.3612   -280.0778
## 19    SavingsAccountBonds.100.to.500  -534.7255   -514.6644
## 20   SavingsAccountBonds.500.to.1000  -637.4894   -616.7925
## 21           EmploymentDuration.gt.7  -226.1536   -213.1417
## 22              Personal.Male.Single   490.3753    501.9253
## 23 OtherDebtorsGuarantors.CoApplicant   604.7289    634.8534
## 24               Property.RealEstate  -713.6104   -691.5967
## 25                Property.Insurance  -450.7156   -427.2650
## 26                 Property.CarOther  -450.7743   -427.9176
## 27           Job.UnemployedUnskilled -1718.7016  -1646.4382
```

```
## 28              Job.UnskilledResident -1254.3615  -1228.0344
## 29                Job.SkilledEmployee -1300.4806  -1276.5771
```

```
ci[,2:3] <- ci[,2:3]*(0.632^0.5)
print("Confidence Interval Scaled")
```

```
## [1] "Confidence Interval Scaled"
```

```
ci
```

```
##                                 Term  lower.2.5. upper.97.5.
## 1                          Intercept  5512.52584  5616.88942
## 2                           Duration    99.26099    99.81435
## 3          InstallmentRatePercentage  -640.09073  -635.35935
## 4                          Telephone  -400.00885  -389.47958
## 5                              Class  -286.26738  -273.24393
## 6        CheckingAccountStatus.lt.0  -234.82629  -222.91070
## 7       CheckingAccountStatus.gt.200  -545.09927  -531.96001
## 8     CreditHistory.NoCredit.AllPaid   649.07737   681.43539
## 9                     Purpose.NewCar -1421.67900 -1320.85876
## 10                   Purpose.UsedCar  -905.95479  -803.11044
## 11       Purpose.Furniture.Equipment -1458.87374 -1357.23025
## 12          Purpose.Radio.Television -1652.37990 -1550.88904
## 13         Purpose.DomesticAppliance -1931.26094 -1827.20816
## 14                   Purpose.Repairs -1361.49650 -1255.16205
## 15                 Purpose.Education -1540.73099 -1438.97290
## 16                Purpose.Retraining -1790.93006 -1677.78269
## 17                  Purpose.Business -1570.85290 -1468.81840
## 18        SavingsAccountBonds.lt.100  -234.01254  -222.65745
## 19    SavingsAccountBonds.100.to.500  -425.09839  -409.15014
## 20   SavingsAccountBonds.500.to.1000  -506.79402  -490.34031
## 21           EmploymentDuration.gt.7  -179.78856  -169.44429
## 22              Personal.Male.Single   389.84062   399.02268
## 23 OtherDebtorsGuarantors.CoApplicant   480.74997   504.69845
## 24               Property.RealEstate  -567.30907  -549.80848
## 25                Property.Insurance  -358.31185  -339.66894
## 26                 Property.CarOther  -358.35845  -340.18777
## 27           Job.UnemployedUnskilled -1366.34074 -1308.89247
## 28             Job.UnskilledResident  -997.19769  -976.26805
## 29               Job.SkilledEmployee -1033.86160 -1014.85876
```

```
print("Single full model CI")
```

```
## [1] "Single full model CI"
```

```
confint(fit)
```

```
##                                       2.5 %      97.5 %
## (Intercept)                        5803.58724 8249.951869
## Duration                            114.73619  135.680095
## InstallmentRatePercentage          -907.67378 -697.801413
## Telephone                          -756.73499 -242.912979
## ClassGood                          -625.42680  -81.235262
## CheckingAccountStatus.lt.0         -560.86038   -9.488975
## CheckingAccountStatus.gt.200      -1143.31044 -201.202787
## CreditHistory.NoCredit.AllPaid      235.19487 1423.776454
## Purpose.NewCar                    -2828.29703 -654.914413
```

```
## Purpose.UsedCar                    -2191.30159   22.986936
## Purpose.Furniture.Equipment        -2881.59508 -685.888865
## Purpose.Radio.Television           -3122.73763 -942.524201
## Purpose.DomesticAppliance          -3879.41510 -897.923124
## Purpose.Repairs                    -2974.12577 -353.411953
## Purpose.Education                  -3067.96844 -705.935812
## Purpose.Retraining                 -3812.34708 -598.708080
## Purpose.Business                   -3044.49769 -809.364713
## SavingsAccountBonds.lt.100          -575.96621    1.165596
## SavingsAccountBonds.100.to.500      -945.53363  -89.503887
## SavingsAccountBonds.500.to.1000    -1130.32068 -115.982230
## EmploymentDuration.gt.7             -489.36764   52.129975
## Personal.Male.Single                 251.22561  724.701134
## OtherDebtorsGuarantors.CoApplicant    39.65711 1197.078236
## Property.RealEstate                -1104.65760 -297.335228
## Property.Insurance                  -838.28281  -51.274726
## Property.CarOther                   -808.26802  -80.111262
## Job.UnemployedUnskilled            -2528.98004 -838.453331
## Job.UnskilledResident              -1691.97405 -798.404704
## Job.SkilledEmployee                -1649.85593 -931.881578
```

These CIs are tighter than single model CIs. # • Summarize results.

1. Using step-wise regression model I've first selected 28 predictor variables.
2. After that I did train test split and calculated all coefficients, holdout R2 and % fall in R2.
3. From the coefficients histogram it was found that all the coefficients value satisfy the assumption of central limit theorem.
4. Confidence interval of training and test data set were narrower than single sample confidence interval.